458.1

UNITED STATES DEPARTMENT OF AGRICULTURE

BUREAU OF ENTOMOLOGY AND PLANT QUARANTINE

FOREST INSECT INVESTIGATIONS

MULTIPLE CORRELATION ANALYSIS OF GROSS INSECT LOSS TO HAZARD FACTORS ON KLAMATH BASIN PLOTS, OREGON

By
F. P. Keen
Senior Entomologist
Bureau of Entomology and Plant Quarantine

Forest Insect Laboratory
445 U. S. Court House
Portland, Oregon
June 20, 1939

MULTIPLE CORRELATION ANALYSIS OF
GROSS INSECT LOSS TO HAZARD FACTORS
ON
KLAMATH BASIN PLOTS, OREGON

By
F. P. Keen
Senior Entomologist
Bureau of Entomology and Plant Quarantine

Forest Insect Laboratory
445 U. S. Court House
Portland, Oregon
June 20, 1939

Introduction.

For many years investigations have been conducted in ponderosa pine forests of the Pacific slope to determine the underlying or contributing causes of bark-beetle outbreaks. Some of the questions for which answers have been sought are:

(1) To what extent is beetle damage dependent upon forest or environmental conditions?

(2) What forest conditions represent a hazard? and

(3) What is the relative weight or importance of the various hazard factors?

In the Klamath Basin of southern Oregon, a large number of full-section (640-acre) sample plots were first established in 1921, in connection with the Southern Oregon-Northern California Pine Beetle Control Project. Each year these plots have been cruised and the current annual beetle loss tabulated. In 1932, a 10-percent cruise was made of the green stand and in later surveys environmental factors, such as site, elevation, forest type, growth rate, and plant associations were determined for each plot and for various parts of each plot.

An analysis of these basic records should throw some light on the above questions. And so, with 10 years of records (1921 to 1930, inclusive) on 25 full-section sample plots as a base, a multiple correlation analysis was made, using such environmental factors as were measurable and could be expressed in numerical values. The method of analysis outlined by Wallace and Snedecor*

*Wallace, H. A., and Snedecor, G. W. "Correlation and Machine Calculation" - Iowa State College of Agr. & Mech. Arts., Vol. XXX, No. 4, June 24, 1931.

was followed. Field Aide R. C. Grant assisted with the computations, especially in solving the normal equations and in handling the partial correlations.

<u>The Basic Plot Data.</u>

The factors included in this analysis were as follows:

<u>Dependent Variable:</u>

y = Average annual gross loss in terms of percent of volume of stand killed. This figure was obtained directly from the annual beetle survey record. It is the average volume loss for the period 1921-1930 inclusive x 100, divided by the original volume of the stand as of 1921.

<u>Independent Variables:</u>

x_1 = Site index. This is the height of dominant trees at age 100 years and was obtained from height curves prepared from Abney measurements on each plot.

x_2 = Volume per acre in Mbm (as of 1921). A 10-percent cruise was made on each plot in 1932, and by correcting for growth and losses, the original 1921 stand was computed.

x_3 = Growth reduction percent. This is the percent by which the growth for 1921-1930 dropped below the normal rate established by previous decades. It was determined by the measurement of 20 or more increment cores taken

from Class 3A trees on each plot. The combined growth rate of all cores from all plots showed a normal rate of 19.0 mm. per decade, which was used as the base rate.

x_4 = Volume of normal growth in board feet. This value is estimated volume of growth uncorrected for reduction, obtained by use of an alignment chart prepared by F. A. Briegleb of the Pacific Northwest Forest Experiment Station. The alignment chart combines site index, stand volume, proportion of Dunning's Classes 1, 2, 3, and 6 in the stand to estimate normal growth. (The average of the period 1900 to 1935 was assumed to be normal in preparing the chart.)

It would have been desirable to have included stand structure or the percent of susceptible trees in the analysis. Unfortunately, no classification of the green stand was available until 1932, after the losses under consideration (1921-1930) had taken place. Since there was no way to reconstruct the stand structure as of 1921, and since it is obvious that the percent of susceptible trees present in 1932 could have little relation to the losses which occurred previously, this factor was not included. It is planned to include this factor in another analysis now under way covering all plots in eastern Oregon and Washington and losses subsequent to 1930.

TABLE NO. I

Average Gross Annual Loss of Ponderosa Pine
and
Four Associated Variables
on Twenty-five 640-Acre Sample Plots
Klamath Basin, Oregon
Period 1921-1930

Plot Description				y	X_1	X_2	X_3	X_4
Name	T.	R.	Sec.	Annual Loss % Stand	Site Index	Volume per Acre Mbm	Growth Reduction Percent	Volume Normal Growth Bd. ft.
Merritt Creek	33S	14E	34	.79	85	22	-26	109
Crowder Flat	47N	12E	4	.87	80	16	-19	87
[illegible]llowe	39S	13E	5	.91	73	15	-30	89
Owens	38S	15E	1	.91	85	21	-33	129
Aspen Lake	37S	7E	34	1.11	88	16	-19	130
Eagle Ridge	37S	7E	16	1.12	84	18	-25	94
Johnson Prairie	39S	5E	7	1.23	100	18	-30	141
Deming Creek	36S	15E	25	1.45	80	15	-43	87
Royston	38S	12E	10	1.53	82	13	-10	76
Meryl Creek	35S	15E	20	1.58	80	21	-23	128
Quartz Valley	38S	17E	7,8	1.75	80	18	-22	115
Meryl Creek	35S	14E	11	1.89	80	21	-32	104
Clover Sta.	38S	6E	36	2.19	84	20	-36	101
Antelope	36S	8E	28	2.22	85	12	-30	72
Jenny Creek	40S	4E	34	2.26	80	9	-28	46
Pokegama	40S	5E	36	2.63	78	12	-24	54
Crowder Flat	47N	11E	18	2.63	78	14	-39	100
[illegible]ber	38S	13E	24	2.97	80	11	-22	70
[illegible]ing Creek	36S	15E	8	3.02	73	11	-28	48
Sycan	34S	12E	2	3.14	78	18	-43	72
Round Lake	39S	8E	5	3.34	74	12	-29	50
Willow Flat	37S	14E	20	3.40	75	13	-39	83
Whitworth	37S	16E	17	3.74	75	12	-55	55
Ferguson	35S	13E	33	4.06	70	11	-42	47
Klamath Canyon	41S	6E	9	5.94	70	6	-44	20
Totals				56.68	1,997	375	-751	2,107
Means				2.2672	79.88	15.00	-30.04	84.28

In Table I the plots included in the multiple correlation analysis are shown, together with the basic numerical values for each of the variables used.

Simple Correlations.

The first step involved the computation of the simple correlation coefficients between the various factors. These computations showed the following relationship:

		Coefficient of Correlation	Significance
y - (% of stand killed) correlates with:			
x_1 - Site index	ryx_1 =	-.65	High
x_2 - Volume pine	ryx_2 =	-.70	High
x_3 - Growth reduction	ryx_3 =	+.64	High
x_4 - Growth rate	ryx_4 =	-.76	High
x_1 - (site index) correlates with:			
x_2 - Volume of pine	rx_1x_2 =	+.54	High
x_3 - Growth reduction	rx_1x_3 =	-.38	N.S.*
x_4 - Growth rate	rx_1x_4 =	+.74	High
y - % of stand killed	rx_1y =	-.65	High
x_2 - (volume of pine) correlates with:			
x_3 - Growth reduction	rx_2x_3 =	-.20	N.S.
x_4 - Growth rate	rx_2x_4 =	+.84	High
x_1 - Site index	rx_2x_1 =	+.54	High
y - Stand killed	rx_2y =	-.70	High

* Not significant.

Simple Correlations (continued)		Coefficient of Correlation	Significance
x_3 - (Growth reduction) correlates with:			
x_4 - Growth rate	rx_3x_4 =	-.33	N.S.
x_1 - Site index	rx_3x_1 =	-.38	N.S.
x_2 - Volume pine	rx_3x_2 =	-.20	N.S.
y - Losses	rx_3y =	+.64	High
x_4 - (Growth rate - normal) correlates with:			
x_1 - Site index	rx_4x_1 =	+.74	High
x_2 - Volume pine	rx_4x_2 =	+.84	High
x_3 - Growth reduction	rx_4x_3 =	-.33	N.S.
y - Losses	rx_4y =	-.76	High

It is interesting to examine these relationships in some detail. In the first set it will be noted that losses are significantly correlated with all four independent variables, but most strongly with growth rate. Computed growth rate obviously is so closely associated with site index and volume of pine that these three variables always move together. As they go up, losses go down.

On the other hand, growth reduction from this analysis shows no significant correlation with either growth rate, site index, or volume of pine. (It might if a larger sample were available.) But it does show a highly significant relationship to losses. Thus growth may be suddenly reduced for one reason or another on good sites as well as poor, or in forests where the normal growth

rate is high. That this happens in the field is borne out by experience. Frequently trees growing under ample moisture conditions are more quickly injured by a drought than those growing on dry rocky hillsides where they are adjusted to limited moisture conditions. Thus it is the change in growth rate that is important rather than its relative normal rate.

<u>The Normal Equations.</u>

The next step involved the solving of four normal equations and the derivation of the four regression coefficients in the regression equations of the form

$$\bar{y} = My + \beta_{yx_1} \frac{\sqrt{\Sigma y^2 - \Sigma y\,(My)}}{\sqrt{\Sigma x_1^2 - (\Sigma x_1)\,Mx_1}} (x_1 - Mx_1) + \beta_{yx_2} \text{ etc.}$$

with the following results:

$$\bar{y} = 226.72 + (-.1232)\,\frac{608.63}{31.16}\,(x_1 - 79.88)$$

$$+ (-.3692)\,\frac{608.63}{20.74}\,(x_2 - 15.00)$$

$$+ (+.4539)\,\frac{608.63}{47.95}\,(x_3 - 30.04)$$

$$+ (-.2110)\,\frac{608.63}{153.78}\,(x_4 - 84.28)$$

which gave the following regression equation:

$$\bar{y} = -2.5040x_1 - 10.8344\,x_2 + 5.7613\,x_3 - .8350\,x_4 + 486.5599$$

TABLE II

Annual Loss Estimated from Regression Equation

Observation	% of Stand Killed y	Estimated % of Stand Killed $\bar{y}$	Error of Estimate $y - \bar{y}$	d^2
1	.79	.94	.15	225
2	.87	1.50	-.63	3969
3	.91	2.40	-1.49	22201
4	.91	1.29	-.38	1444
5	1.11	.94	.17	289
6	1.12	1.47	-.35	1225
7	1.23	.96	.27	729
8	1.45	1.84	-.39	1521
9	1.53	1.34	.19	361
10	1.58	.84	.72	5184
11	1.75	1.22	.53	2809
12	1.89	1.56	.33	1089
13	2.19	1.83	.36	1296
14	2.22	2.56	-.34	1156
15	2.26	3.12	-.86	7396
16	2.63	2.54	.09	81
17	2.63	2.81	-.18	324
18	2.97	2.35	.62	3844
19	3.02	3.06	-.04	16
20	3.14	2.84	.30	900
21	3.34	2.96	.38	1444
22	3.40	3.13	.27	729
23	3.74	4.40	-.66	4356
24	4.06	3.95	.11	121
25	5.94	4.83	1.11	12321
			Check	7.5030
			-5.47	
			5.45	

$$dy_{x_1x_2x_3x_4} = \sqrt{\frac{7.5030}{25-5}} = .61\%$$

By substituting the values of x_1, x_2, x_3, and x_4 as given in table I in the regression equation on page 7, it is possible to calculate the estimated percent of insect loss and errors of estimate. These are shown in table II. The range in average annual loss on the plots was from .79 to 5.94 percent of the stand, with an average of 2.267. The standard error of estimate from table II was found to be ± .61 percent, which is rather high.

On examining this table, we find that the multiple regression equation tends to overestimate the losses on plots where losses were light (less than 1 percent per year), and underestimates losses at the other end of the table on plots where losses were extremely heavy (4 percent per year and over). Thus the four independent variables included in this analysis fail to explain some of the extreme losses or reasons why some plots have shown such light losses. One important reason for this is the fact that the regression equation assumes a straight line relationship between losses and the other variables, when it is more reasonable to suppose that some of these relationships may be curvilinear in character. For instance, one would not expect growth rate to influence loss at any fixed rate, but rather at an accelerating rate as growth declines. (In the regression equation on page 7, with zero growth loss would equal 2.96 percent per year.) Because of the limited amount of data and the wide scatter of points on plotting, no attempt was made to introduce curvilinear regression in this analysis, which was essentially an exploratory search for important hazard factors.

Plot number 3, which is badly overestimated by the formula, is particularly worthy of study. This section on Goodlowe Mountain (T. 39S, R. 13E, Sec. 5) in spite of poor site, proximity to the desert edge, and poor growth rate, has consistently shown very light losses. Why this has been so is hard to explain. Goodlowe Mountain is somewhat isolated from other pine timber in the Bonanza Area of the Klamath Basin, and although included within the boundaries of the Southern Oregon-Northern California Pine Beetle Control Project, was never disturbed with control work. Neighboring areas of exactly the same type, site, and elevation have suffered losses of from 15 to 34 percent of the stand in the 10-year period 1921 to 1930, yet on Goodlowe only 9 percent of the stand was killed during this period. Just why this difference, has been the subject of considerable speculation, but the answer has not been found.

<u>The Coefficient of Multiple Correlation.</u>

The coefficient of multiple correlation, calculated from the formula

$$R^2 = \beta_{yx_1} r_{x_1y} + \beta_{yx_2} r_{x_2y} + \beta_{yx_3} r_{x_3y} + \beta_{yx_4} r_{x_4y}$$

was found to be

$$R^2 = (-.1282)(-.6532) + (-.3692)(-.7054) + (.4539)(.6444) + (-.2110)(-.7670)$$

$$= .7984$$

$$R = \sqrt{.7984} = .8953$$

This coefficient of correlation expresses the relationship between the estimated and actual value of losses as shown in table II. Since a highly significant value of R for 25-5 degrees of freedom is .685, we may conclude that this multiple correlation coefficient indicates a highly significant association between the five variables. The regression equation may be said to explain approximately 80 percent of the variation in losses on the different plots used in this analysis and for the time under consideration (1921-1930). In other words, given the insect population conditions which existed in this area during the period 1921-1930, 80 percent of the variation in losses between the different plots of the area can be explained by these four factors, i.e., site index, volume of stand, normal growth rate, and percent of reduction in growth.

<u>The Standard Error of Estimate.</u>

The portions of the total sum of squares due to regression and not accounted for may be derived from the formulae:

$$R^2 (\Sigma y^2 - (\Sigma y)\ My) \text{ and } (1-R^2)(\Sigma y^2 - (\Sigma y)\ My)$$

The results are shown in table III and indicate that the standard error of estimate is 0.61 percent, which is essentially the same as derived in table II by simply squaring the errors of estimate and dividing by the degrees of freedom.

TABLE III

Analysis of Variation in Multiple Regression
Five Variables

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square or Variance	Standard Deviation
Due to regression	4	29.5754	7.3938	
Not accounted for	20	7.4679	.3734	.61%
Total	24	37.0433	1.5435	1.24%

Test of Significance.

By calculating the standard deviation of the betas in the regression equation on page 7, we can apply the "t" test and determine the significance and weight of each of the factors used in this analysis. The results of these calculations gave the following:

TABLE IV

Test of Significance

Variable	Symbol	Standard Regression Coefficient " "	Standard Deviations of the 's	"t" Ratio	Percent or Weight
Site	x_1	-.1282	.1546	.83	11
Volume pine per acre	x_2	-.3692	.1934	1.91	32
Growth reduction %	x_3	+.4539	.1095	4.14	39
Annual growth bd. ft.	x_4	-.2110	.2410	.88	18
Total		1.1623			100

t = 2.086 for least significance
t = 2.845 for high significance

Since in the "t" test the ratio of each β divided by its respective standard deviation should be 2.026 for least significance and 2.845 for high significance, it is evident from the above that growth reduction is highly significant, volume per acre just below the limit of significance, and neither site nor annual growth significant.

At first thought this is most surprising. We have been thinking of site and growth rate as highly important in explaining variation in losses, and by this test they each fail to show significance.

The reason for the failure of site index to show significance is partly due to the lack of spread in this case. Nearly all the plots ran to site quality IV and the total range in site index was only from 70 to 100. It is unreasonable to expect a variable which is nearly constant to explain any large part of the wide variation in losses. In the second place, site is at its optimum in the middle of the pine belt and decreases at both higher and lower elevations. Beetle losses increase at the lower elevations, but not at the higher, and hence do not move together with changes in site.

Annual growth rate and losses were shown by simple correlation to have a highly significant coefficient of correlation (-.76), but in the multiple regression equation its role is unimportant. This appears to be a paradox but is not necessarily so. It means that the factor of annual growth alone does not contribute much to

improving the relationship between losses and environmental factors (as expressed by the multiple regression equation), but is highly associated with losses because of its close concurrent association with site, volume of pine, or other important factors.

To summarize: Losses are highly correlated with all four factors used in this analysis, i.e., site index, volume of pine, normal growth rate, and growth reduction. But of these, only growth reduction is highly significant and useful in explaining the variation in losses. The other factors, while related to loss, are only indirectly associated with it in some manner which as yet we cannot explain.

Partial Correlation.

Fourth order partial correlation coefficients were computed to determine what degree of association existed between loss and each of the four independent variables, independent of their common association with each other, with the following results:

Variable	Partial Correlation Coefficient	Significance
Site index x_1	$rx_1y \cdot x_2x_3x_4 = -.147$	None
Volume per acre x_2	$rx_2y \cdot x_1x_3x_4 = -.382$	None
Growth reduction % x_3	$rx_3y \cdot x_1x_2x_4 = +.683$	High
Annual growth x_4	$rx_4y \cdot x_1x_2x_3 = -.242$	None

For 20 degrees of freedom (25-5) we find

$r = .423$ least significant

$r = .537$ highly significant.

Therefore, we again find that only growth reduction is highly significant and none of the other factors are significant, although volume per acre is only slightly below the limits. This is very illuminating.

In simple correlation we found site index, volume per acre, and annual growth all highly significant. But the partial correlation coefficients show none of them to be significant when divorced from their common association with each other. They all stand together or fall apart separately, and this is only reasonable when we consider how closely associated annual growth is with site and volume per acre.

If permanent environmental conditions such as slope elevation, site, etc., were always more hazardous to beetle loss than other conditions, we would expect losses to always hit hardest in those spots until the timber stands were completely eliminated. We know that this does not happen. So it seems more reasonable to expect losses, which are variable, to be associated with environmental factors, which are also variable—for instance, soil moisture, stand density, fluctuations in tree growth and vigor, and temperature conditions.

On examining the factors used in our analysis, we find that site is nearly a permanent condition; stand volume changes but slowly; and normal annual growth, being computed from these two factors plus stand structure, also changes slowly. Therefore, it is

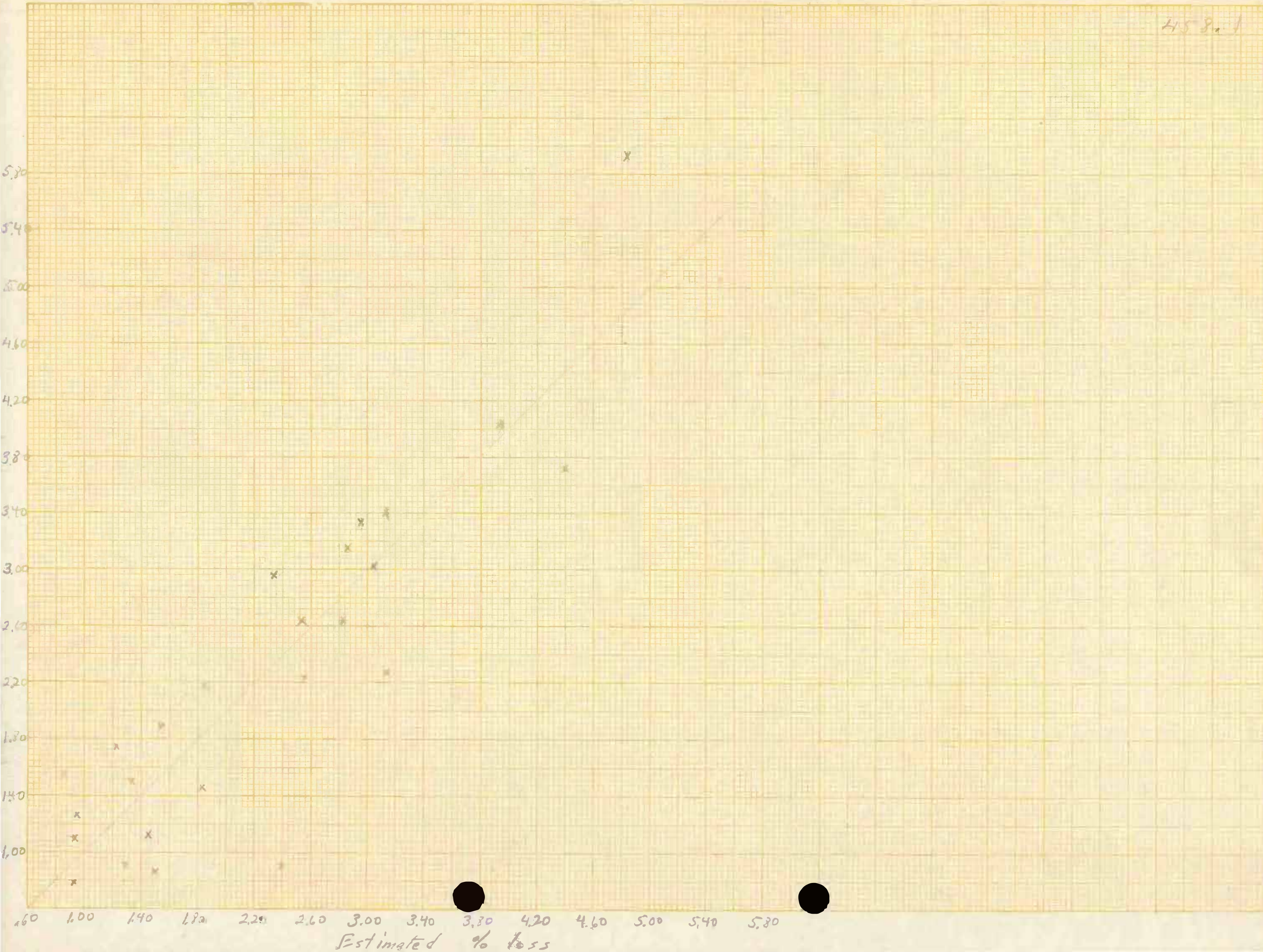
hardly reasonable to expect these more or less constant factors to explain any large part of fluctuating losses. The partial correlation analysis shows that they do not. On the other hand, growth reduction is a fluctuating factor, depending upon seasonal changes. As growth drops, losses increase; and the two fluctuate together. That this is a highly significant factor is not only fully demonstrated by this analysis, but it is reasonable to conclude that it should be important, considering everything we know about western pine beetle behavior and the type of trees and stands it selects for its activities.

Conclusion.

This multiple correlation analysis of four independent variables associated with beetle losses on 25 sample plots in the Klamath Basin has shown that reduction in tree growth is the one outstanding factor indicating hazard.

Site, volume of stand, and normal growth rate, while jointly associated with some variable or variables responsible for fluctuations in loss, do not in themselves explain any significant part of this fluctuation. This failure to show significance may be due either to a real lack of any important relationship between the variables or to an inadequate amount of data in this analysis. However, because these three factors are indirectly associated with variables responsible for insect losses, as indicated by their significant simple correlation coefficients, they have some value in indicating the probable location of hazardous areas.

Measured over Estimated   Source: Table II Station Report 1939



Leave [illegible]

TABLE II

Calculation of Correlation Coefficients *

	Site Index x_1	Volume per Acre x_2	% Normal Growth x_3	Annual Growth x_4	% Loss $y \times 100$	S
Sums	1997.	375.	751.	2107.	5668.	10898.
Means	79.88	15.00	30.04	84.28	226.72	435.92
x_{1-1}	160491	30304	59419	171831	440373	862418
x_{1-2}	159520	29955	59990	168307	452760	870532
x_{1-3}	971	349	-571	3524	-12387	-8114
x_{1-4}	31.16	646.26	1494.12	4791.78	18964.91	
x_{2-1}	....	6055	11070	34295	76116	157840
x_{2-2}	....	5625	11265	31605	85020	163470
x_{2-3}	....	430	-195	2690	-8904	-5630
x_{2-4}	....	20.74	994.48	3189.40	12622.99	
x_{3-1}	....	....	24859	60853	189073	345274
x_{3-2}	....	....	22560	63294	170267	327376
x_{3-3}	....	....	2299	-2441	18806	17898
x_{3-4}	....	....	47.95	7373.75	29183.81	
x_{4-1}	....	....	....	201227	405914	874120
x_{4-2}	....	....	....	177578	477699	918483
x_{4-3}	....	....	....	23649	-71785	-44363
x_{4-4}	....	....	....	153.78	93595.12	
y_1	....	....	....	....	1655482	2766958
y_2	....	....	....	....	1285049	2470795
y_3	....	....	....	....	370433	296163
y_4	....	....	....	....	608.63	

* Calculations according to table 7a of Wallace and Snedecor.

Leave out

TABLE III

Solution of Normal Equations*

Block	Line	Site Index x_1	Volume per Acre x_2	% Normal Growth x_3	Annual Growth x_4	% Loss y	S
x_1	1	1.0000	.5400	-.3755	.7354	-.6532	1.2467
	2	....	-.5400	.3755	-.7354	.6532	
x_2	3	....	1.0000	-.1961	.8434	-.7054	1.4819
	4	....	-.2916	.2028	-.3971	.3527	-.6732
	5	....	.7084	.0067	.4463	-.3527	.8087
	6	....	-1.0000	-.0094	-.6300	.4978	-1.1415
x_3	7	....		1.0000	-.3310	.6444	.7418
	8	....		-.1410	.2761	-.2453	.4681
	9	....		-.0000	-.0042	.0033	-.0076
	10	....		.8590	-.0591	.4024	1.2023
	11	....		-1.0000	.0688	-.4684	-1.3996
x_4	12	....			1.0000	-.7670	1.4808
	13	....			-.5408	.4804	-.9168
	14	....			-.2812	.2222	-.5095
	15	....			-.0041	.0277	.0827
	16	....			.1739	-.0367	.1372
	17	....			-1.0000	.2110	-.7889
β_{yx_4}	1	....			-.2110	-.2110	
β_{yx_3}	2	....		.4539	-.0145	.4684	
β_{yx_2}	3	....	-.3692	-.0043	.1329	-.4978	
β_{yx_1}	4	-.1282	.1994	.1704	.1552	-.6532	

* Calculations according to table 8a of Wallace and Snedecor.

Use of the regression in estimating annual loss %.

Equation:

$$\bar{Y} = -2.5040\,X_1 - 10.8344\,X_2 + 5.7613\,X_3 - .8350\,X_4 + 486.5599$$

where :- X_1 = site index ; X_2 = Vol./Acre Mbm. ;
X_3 = Growth Reduction Percent ; X_4 = Vol. Normal Growth B.F.

By substituting the means of each independent variable the $\bar{Y}$ estimate is equal to the mean of the dependent variable.

$$\bar{Y} = -2.5040\,(79.88) - 10.8344\,(15.00) + 5.7613\,(30.04) - .8350\,(84.28) + 486.5599$$

$+ 5.7613\,X_3$

$$\bar{Y} = -200.0195 - 162.5160 + 173.0694 - 70.3738 + 486.5599$$

$\bar{Y} = 2.2672$ estimate $Y = 2.2672$ actual mean.

The growth reduction percent should be substituted as a positive number if it is actually negative; if it is actually positive, 100 should be added to it.

Use of the multiple regression equation in estimating annual loss % from growth reduction.

y = Mean annual gross loss in terms of percent of volume killed

x_3 = growth reduction percent.

Combined radial growth from all 3 A cores from all plots was 19.0 mm per decade. This was used as the base rate or normal for previous decade 1900-1919.

$$\bar{y} = -200.0195 - 162.5160 + 5.7613\, x_3 - 70.3738 + 486.5599$$

$$\bar{y} = \frac{5.7613 x_3 + 53.6506}{100}$$

Test.

30.04 % mean an growth reduction

$$\bar{y} = \frac{5.7613(30.04) + 53.6506}{100}$$

$\bar{y}$ = 2.267% or mean annual loss.

Reg.

$\bar{Y} = -2.043(88) - 9.523(16) + 5.811(-19) - 1.049(130) + 446.66$

$\bar{Y} = -180.00 - 152.40 + .110.40 - 135.20 + 446.66$

180.00
152.40
135.20
467.60

110.40
446.66
557.06
467.60
89.46

Aspen Lake Sample.
.89.46 % est regression
1.11 actual

Original equation.

$\bar{Y} = -2.043(80) - 9.523(15) + 5.811(-30.04) - 1.049(84.28) + 446.66$

$- 163.80 - 143.00 + 174.50 - 88.40 + 446.66$

163.80
143.00
88.40
395.20

446.66
174.50
621.16
395.20
225.96

Check:

2.2672 actual average.
2.26. est. by regression equ.

New equation.

$\bar{Y} = -2.5040 X_1 - 10.8344 X_2 + 5.7613 X_3 - .8350 X_4 + 486.5599$

$-2.5040(80) - 10.8344(15) + 5.7613(-30.04) - .8350(84.28) + 486.5599$

$-200.00 - 162.30 + 172.00 - 70.30$

200.00
70.30
162.30
432.60

Check:

2.26 est by reg.
2.2672 act. ave.

486.5599
172.000
658.5599
432 60
225.9599

The growth reduction percent was probably handled as a positive number in the problems so in substituting in the regression equation signs are disregarded.

− 200.01952
− 162.5160
+ 173.069452
− 70. 37380
486. 5599

22672

	Reduction	Loss %
-173.069452	-30.04	2.267%
53.6506	0	.53%
~~226.7200~~		
-119.4188	+30	-1.19% growth.

-1.19%

$$\begin{array}{r} 2.267 \\ .53 \\ \hline 1.737 \end{array}$$

30% increase in growth = 76% reduction in loss

$$
\begin{array}{r}
-200.0195 \\
-162.5160 \\
-70.3738 \\
\hline
-432.9093
\end{array}
\qquad
\begin{array}{r}
486.5599 \\
432.9093 \\
\hline
53.6506
\end{array}
$$

$$Y = -200.0195 - 162.5160 + 5.7613\,X_3 - 70.3738 + 486.5599$$

$$= \frac{5.7613\,X_3 + 53.6506}{100}$$

$$y = \cancel{\quad X_3 + 9.3 + 233}$$